(54) Title: IN SILICO SCREENING FOR PHENOTYPE-ASSOCIATED EXPRESSED SEQUENCES

(57) Abstract: The present invention provides methods for determining whether a nucleic acid sequence is a marker for a phenotype or cell type of interest which comprises providing a database of expressed sequence tag sequences (EST's) from the species; placing said EST's in groups termed clusters based on homology of EST's within each cluster; determining for each cluster the total number of EST"s within said cluster; ordering said clusters sequentially based on the number of EST's in each cluster; dividing said ordered clusters into subranges based on the number of EST's per cluster; determining for each cluster subrange obtained from step (e) the number EST's within said cluster which are expressed in said predetermined cell type of interest; calculating according to a normal distribution the number of clusters in each subrange expected to contain a predetermined threshold percentage of EST's expressed in said cell type of interest, wherein said threshold percentage is a percentage from about 10% to about 100%; determining the number of clusters in each subrange observed to contain said predetermined threshold percentage of EST's expressed in said predetermined cell type; and identifying subranges having an observed number of clusters that meet said predetermined threshold percentage greater than the number of clusters expected to meet said predetermined threshold percentage for the subrange according to normal distribution; wherein if the percentage of EST's expressed in said cell type of interest in a cluster identified is equal to or greater than said predetermined threshold percentage, the cluster contains a nucleic acid that is a marker for the cell type of interest.

| | | | | | |
|---|---|---|---|---|---|
| Hs.217766 | POM105 | Ovary carcinoma | | | |
| Hs.217882 | POM106 | glioma, colon carcinoma, kidney tumors, prostate tumors, lung carcinoma, hypernephroma, head and neck carcinoma, duodenal carcinoma, melanoma, pancreatic carcinoma, uterus tumors | | SEQ. ID NO: 336 | SEQ. ID NO: 337 SEQ. ID NO: 339 |
| Hs.220529 | CEACAM5 Carcinoembryonic antigen-related cell adhesion molecule 5 | Pancreas carcinoma, colon carcinoma, stomach carcinoma, head and neck carcinoma, lung carcinoma leiomdioma, breast carcinoma | KNOWN TUMOR MARKER | SEQ. ID NO: 340 | SEQ. ID NO: 341 |
| Hs.222056 | POM107 Homo sapiens cDNA FLJ11572 fis, clone HEMBA1003373 | Stomach carcinoma, head and neck carcinoma,breast carcinoma | | SEQ. ID NO: 342 | SEQ. ID NO: 343 |
| Hs.225083 | POM108 | Melanoma, ovary tumors, colon carcinoma, parathyroid tumor,kidney tumors, head and neck carcinoma | | SEQ. ID NO: 344 | SEQ. ID NO: 345 |
| Hs.227098 | GCMB Glial cells missing homolog b (Drosophila) | parathyroid_tumor | | SEQ. ID NO: 346 | SEQ. ID NO: 347 |
| Hs.239107 | POM109 | Lymphoma, germ cell tumors, head and neck carcinoma | | SEQ. ID NO: 348 | SEQ. ID NO: 349 |
| Hs.239891 | GPR35 G protein-coupled receptor 35 | B-cell chronic lymphocytic leukemia, colon carcinoma, pancreas and carcinoma | SURFACE | SEQ. ID NO: 350 | SEQ. ID NO: 351 |
| Hs.241381 | CRSP7 Cofactor required for Sp1 transcriptional activation, subunit 7 (70kD) | Pancreatic carcinoma, duodenal carcinoma, ovary carcinoma, melanoma, osteosarcoma, glioma, leiomyosarcoma, germ cell tumors | | SEQ. ID NO: 352 | SEQ. ID NO: 353 |
| Hs.241407 | SERPINB13 Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 13 | ORAL carcinoma, cervical carcinoma, head and neck carcinoma | | SEQ. ID NO: 354 | SEQ. ID NO: 355 |
| Hs.243920 | POM110 | Pancreas carcinoma | | SEQ. ID NO: 356 | SEQ. ID NO: 3 |
| Hs.244378 | SLC2A6 Solute carrier family 2 (facilitated glucose transporter), member 6 | Hypernephroma, pancreatic carcinoma, glioma, lung carcinoma, neuroblastoma, renal cell carcinoma, adrenal gland tumors | | SEQ. ID NO: 358 | SEQ. ID NO: 3 |
| Hs.246781 | POM111 | parathyroid_tumor, lung carcinoid tumors, germ cell tumors, hepatocellular carcinoma, stomach carcinoma, breast carcinoma | | SEQ. ID NO: 360 | SEQ. ID NO: 3 |